

We claim as our invention:

1. A market data acquisition system, comprising:

a means for retrieving event and embedded content data from a plurality of set-top  
boxes;

a means for retrieving content attributes from a content attribute database;

a means for correlating retrieved set-top box event data with content attributes to  
produce data indicating which content was experienced through the plurality of  
set-top boxes;

a means for retrieving demographic information from a demographic information  
database; and

a means for correlating demographic information to data indicating which content  
was experienced through the plurality of set-top boxes to produce, in response to a  
query, data indicating content experienced by a demographic group or set of  
demographic groups.

2. The market data acquisition system of Claim 1, in which said state-change data  
collection means collects data from said set-top boxes without access to set-top box  
specific personal or demographic information, thereby providing a layer of privacy  
to set-top box assignees.

3. The market data acquisition system of Claim 2, in which set-top box specific  
demographic or other personal data may be collected when requested or with  
approval given by a set-top box assignee, governmental agency, or other such  
authority.

4. The market data acquisition system of Claim 3, in which a list of set-top box  
identification numbers and zip codes or other geographic identifiers corresponding to  
set-top box installation points is provided to the present invention for each set-top  
box.

5. The market data acquisition system of Claim 1, in which said content attribute database  
is maintained as part of the system.

6. The market data acquisition system of Claim 1, in which said content presentation  
system is maintained external to the present invention.

7. The market data acquisition system of Claim 1, in which said demographic information database is maintained.
8. The market data acquisition system of Claim 1, in which said demographic information database is maintained externally.
- 5 9. The market data acquisition system of Claim 1, in which said queries are entered through a graphical, command-line, or natural language interface.
10. The market data acquisition system of Claim 9, in which said queries can result in the generation reports for any time segment or set of time segments with high precision.
11. The market data acquisition system of Claim 9, in which said queries result in the generation of reports generated individual content or for a set of content.
- 10 12. The market data acquisition system of Claim 9, in which said queries result in generation of said reports for persons fitting a demographic specification, persons fitting a demographic category, or persons fitting sets of demographic specifications and demographic categories.
- 15 13. The market data acquisition system of Claim 9, in which said queries result in reports generated for specific behaviors.
14. The market data acquisition system of Claim 9, in which said queries include one or more highly-specific times, demographic specifications, viewer behaviors, and content descriptions.
- 20 15. The market data acquisition system of Claim 9, in which said results are presented in a graphical manner, such as through a pie chart or bar graph.
16. The market data acquisition system of Claim 9, in which said results are presented as a spreadsheet or other grid.
17. The market data acquisition system of Claim 9, in which said results are presented as natural language.
- 25 18. The market data acquisition system of Claim 1, in which said content information is obtained from a source external to the present invention.

19. The market data acquisition system of Claim 1, in which said content information is embedded in content as it is presented to a set-top box.

20. A method of correlating dynamic and static datasets sharing at least one common characteristic and having an assumed relationship, and using such correlations to determine rule systems between the sets, comprising the steps of:

- selecting subsets of said datasets sharing a common characteristic;
- expressing the assumed relationship as a mathematical assumption;
- defining an error function which describes the two datasets in terms of said mathematical assumption;
- performing fitting procedures to account for errors in the assumed relationship; and
- performing fitting procedures which account for errors in the definition of the common subsets.

21. The method of Claim 20, in which said dynamic data corresponds to set-top box event data.

22. The method of Claim 21, in which said static data corresponds to demographic data.

23. The method of Claim 22, in which correlations are drawn between set-top box event data and demographic to determine the relationship of demographics to content viewership.

24. A method of testing assumptions pertaining to relationships between two disparate datasets sharing at least one common aspect, comprising the steps of:

- entering such assumptions through a user interface;
- selecting sample data from a first dataset;
- determining correlations between said selected data and data stored in a second dataset; and
- establishing assumption validity based on such correlations.

25. A method of determining individual characteristics by correlating dynamic and static datasets sharing at least one common characteristic and having an assumed relationship, comprising the steps of:

- selecting subsets of said datasets sharing a common characteristic;
- expressing the assumed relationship as a mathematical assumption;

defining an error function which describes the two datasets in terms of said mathematical assumption;

performing fitting procedures to account for errors in the assumed relationship;

storing such correlations in an individual-specific array; and

iteratively repeating this process.

26. The method of Claim 25, in which said dynamic dataset corresponds to set-top box data.

27. The method of Claim 26, in which said static dataset corresponds to demographic data.

28. The method of Claim 27, in which said individual-specific data corresponds to a set-top box identification number or other privacy-compliant identification number.

29. The method of Claim 28, in which an IDM algorithm determines said correlations.

30. A method of dynamically determining the demographic identity of an individual operating a set-top box, comprising the steps of:

monitoring set-top box events for a plurality of set-top boxes;

correlating set-top box events with demographic characteristics;

applying IDM calculation techniques to determine probabilities for demographic characteristic and set-top box event dataset correlations;

ascribing demographic characteristic probabilities to each set-top box over time based on observed set-top box events and their relationship to such IDM probabilities;

evaluating such ascribed demographic characteristic probabilities over time through statistical analysis;

fitting probabilities ascribed to demographic characteristics to statistically determine the most likely set of constant dataset possibilities for each set-top box; and,

fitting set-top box possibility sets to IDM probability sets for a set-top box event.

31. The method for determining the demographic identities of individuals in a home, business, or other location containing a set-top box according to the method of Claim 30, further comprising the steps of:

storing said demographic identities in an array over time; and

applying statistical analyses to said array to determine predominant demographic identities for a given set-top box.

32. A system for directing content to a specific demographic group, comprising:

- an array identifying demographic identities associated with set-top boxes;
- a means for entering a demographic group to be targeted;
- a means for entering the content, or a reference to such content, to be directed to a demographic group;
- a means for entering times and other properties indicating a preferred content delivery method; and
- a means of delivering content to a set-top box corresponding to requested demographic information.

33. The system of Claim 32 in which said content refers to advertising.

34. A system for directing content to set-top boxes exhibiting a behavior or pattern of behaviors when a specified content type is presented, comprising:

- a set of set-top box events with specific time recordings for each event;
- a set of content properties;
- a means for correlating set-top box events to content properties;
- a means for entering desired set-top box event/content property correlations;
- a means for delivering content to those set-top boxes corresponding to said set-top box event/content property correlations.

35. The system of Claim 34 in which said content refers to advertising.

36. A method of determining the effect of content attributes on content ratings, comprising the steps of:

- obtaining content attributes from embedded content information or from external sources;
- recording set-top box events as content is experienced;
- correlating set-top box events to content attributes; and,
- analyzing such correlations over time to determine the effect of content attributes on content ratings.

37. The method of Claim 36 in which said content attributes include times at which various content attributes are presented to a set-top box, thereby allowing the present invention to provide detailed correlations between such attributes and set-top box events.

38. A method of determining the effect of content attributes on content ratings for a specific demographic group, comprising the steps of:

obtaining content attributes from embedded content information or from external sources;

5 recording set-top box events as content is experienced;

correlating set-top box events to content attributes;

correlating set-top box events and content attributes to demographic characteristics for each set-top box; and

10 analyzing such correlations over time to determine the effect of content attributes on content ratings for specific demographic groups.

39. The method of Claim 38 in which said content attributes include times at which various content attributes are presented to a set-top box, thereby allowing the present invention to provide detailed correlations between set-top box events, set-top box demographics, and content attributes.

15 40. A method of creating new content based on previously experienced content and content ratings, comprising the steps of:

obtaining content attributes from embedded content information or from external sources;

recording set-top box events as content is experienced;

20 correlating set-top box events to content attributes;

analyzing such correlations over time to determine the effect of content attributes on content ratings; and

analyzing the effect of content attribute order on content ratings; and

determining a preferred content attribute set and content attribute presentation order.

25 41. The method of Claim 40 in which said content attributes include times at which various content attributes are presented to a set-top box.

42. A method of creating new content based on previously experienced content and content ratings, where such new content is directed toward a demographic group, comprising the steps of:

30 obtaining content attributes from embedded content information or from external sources;

recording set-top box events as content is experienced;

correlating set-top box events to content attributes;  
 correlating set-top box events and content attributes to demographic characteristics;  
 analyzing such correlations over time to determine the effect of content attributes on  
 content ratings for a given demographic group; and  
 5 analyzing the effect of content attribute order on content ratings for a given  
 demographic group; and  
 determining a preferred content attribute set and content attribute presentation order  
 for a given demographic group.

43. A system for predicting future events based on a proposed dataset, consisting of:

- 10 a dataset of past events;
- a known dataset sharing at least one attribute with said dataset of past events, and  
 with substantially similar attributes to said proposed dataset;
- a means of correlating said dataset of past events with said known dataset to form a  
 new dataset; and,
- 15 a means of correlating said new dataset to said proposed dataset.

44. The system of Claim 43, where said dataset of past events consists of set-top box  
 event data.

45. The system of Claim 44, in which said known dataset consists of sales figures.

46. The system of Claim 44, where said known dataset consists of content attributes and  
 20 content presentation data.

47. The system of Claim 46, where said proposed dataset consists of a set of content  
 attributes for proposed content.

48. A method of predicting future events given a proposed dataset, comprising the steps  
 of:

- 25 monitoring past events;
- correlating said past events with a dataset sharing at least one attribute with said past  
 events, and with a substantially similar structure to the proposed dataset, the  
 results of such are stored in an array;
- correlating said array with said proposed dataset; and
- 30 reporting the results of said array/proposed dataset correlations as a prediction of  
 future events.

49. The method of Claim 48, in which said past events include set-top box events.

50. The method of Claim 49, in which said proposed dataset substantially consists of proposed content attributes.

51. The method of Claim 50, in which said dataset includes previously presented content attributes.

5 52. The method of Claim 51, in which said dataset consists of sales figures.

53. A system of predicting future events for a given demographic segment, comprising:

a dataset of past events;

a demographic dataset sharing at least one attribute with said dataset of past events;

10 a means of correlating said dataset of past events with said demographic dataset and storing the result in an array;

a known dataset sharing at least one attribute with said demographic dataset, and with substantially similar attributes to said proposed dataset;

a means of correlating said array with said known dataset to form a new dataset; and,

15 a means of correlating said new dataset to said proposed dataset.

54. The system of Claim 53 in which said dataset of past events corresponds to set-top box data.

55. The system of Claim 54 in which said demographic dataset shares a zip code or other geographic identifier with said set-top box data.

20 56. The system of Claim 55 in which said known dataset shares a zip code or other geographic identifier with said array.

57. The system of Claim 56 in which said proposed dataset is comprised of proposed content and attributes corresponding thereto.

25 58. A method of predicting future events for a given demographic based on a proposed dataset, comprising the steps of:

monitoring past events;

correlating said past events with a demographic dataset and storing the result in an array;

30 correlating said array with a dataset sharing at least one attribute with said array, and with a substantially similar structure to the proposed dataset, the results of such are stored in an additional dataset;



correlating said additional dataset with said proposed dataset; and  
reporting the results of such correlations as a prediction of future events.

59. The method of Claim 58, in which said past events include set-top box events.

60. The method of Claim 59, in which said demographic dataset and said set-top box  
5 event dataset both contain zip code attributes.

61. The method of Claim 60, in which said proposed dataset substantially consists of  
proposed content attributes.

62. The method of Claim 61, in which said dataset includes previously presented content  
attributes.

10 63. A privacy-compliant data collection and data correlation system comprising:

a means of collecting individual-specific behavior data without knowing individual-  
specific demographic information pertaining to the individual about whom such  
data is collected;

15 a means of accessing demographic data for the region in which the individual  
resides; and

a means of correlating such individual-specific data with such demographic data to  
determine the demographic identity of each individual about whom data is  
collected.

20 64. The privacy-compliant data collection and data correlation system of Claim 63,  
wherein said individual-specific behavior data collection means is comprised of a  
set-top box.

65. A method of predicting behaviors of non-sampled demographic specifications based  
on sampled demographic specifications of a given level comprising the steps of:

25 monitoring past behavior and correlating such behavior with demographic  
characteristics monitored;

breaking a non-sampled demographic specification into sub-specifications for which  
sample data has been collected;

establishing the statistical effects of various rules on each sub-specification and  
those characterizations comprising them; and

30 statistically predicting non-sample behaviors based on such effects.

66. The method Claim 65, further comprising the steps of:

5 observing correlations between behaviors of sampled demographic specifications or sub-specifications and behaviors of non-sampled demographic specifications, and inferring behaviors of such non-sampled demographic specifications from such correlations, such that predicted or observed sampled demographic specification behaviors may be reported as non-sampled demographic specification behaviors within a determinable level of accuracy.

67. A method of reducing the effect of sampling error and sample bias on data

correlations determined between a dynamic dataset and a static dataset based on assumptions about the relationships between such data, such as:

- 10 creating equations to express such assumptions;  
determining error functions which can assist in calculating values for each unknown variable in such equations;  
creating a transformable matrix based on such functions;  
15 inverting said matrix to apply a least-squares approach fitting method to the underlying data;  
normalizing the results of said least-squares fit;  
calculating Pearson-r correlations for such normalized results;  
calculating aspect representation indices for each subset of data within said static dataset;  
20 determining assumption validities for assumptions used as a basis for this process;  
and  
combining said correlations, said aspect representation indices, and said assumption validities to create a set of data correlations and corresponding confidence intervals.

25 68. The method of Claim 67 in which said dynamic dataset represents set-top box event data.

69. The method of Claim 68 in which said static dataset represents demographic information.

30 70. The method of Claim 69 in which the assumption used to relate said set-top box event data with said demographic information is the demographic assumption.

71. The method of Claim 20, in which said fitting procedures include applying additional assumptions to provide missing correlations values.

72. A method of increasing correlation result dataset specificity by reducing possibilities, consisting of the steps:

calculating correlation result dataset characterization values which fall within a predetermined confidence limit using aspect representation indices, inverse demographic matrices, recombination matrices, and specification similarity matrices;

5 creating a matrix of such values for all demographic characterizations for each method used;

utilizing mathematical expressions of the requirement of consistency for distinct value ranges for identical characterizations in the separate matrices, reducing each range for a given characterization to the greatest possible extent within a predetermined confidence interval; thus producing one matrix with one value range for each characterization;

possibly transforming value ranges for all characterizations within said matrix to the same statistical confidence;

15 iteratively reducing all ranges to the greatest possible extent by utilizing both mathematical expressions of the requirement of consistency among all value ranges in said matrix as well as constraints given by actual characterization population numbers; and

adjusting the statistical confidence if necessary to allow for further value range reduction past the point of useful iteration at a previous statistical confidence.

20 73. The method of Claim 72 in which said dataset correlations result from correlations of set-top box event data and demographic data.

74. The method of Claim 72 in which said dataset correlations result from correlations of demographic data and sales data.

25 75. The method of Claim 72 in which said dataset correlations result from correlations of set-top box data and sales data.

76. A method of fitting by convergence and similarity between a static dataset and a dynamic dataset, comprising the steps of:

defining subsets of each dataset;

determining correlations between such datasets;

30 performing a time-based analysis of group representations and additional correlations within said correlations;

assigning weights to such representations and additional correlations; and,

applying such weights and values to determine undefined correlation dataset values.

77. The method of Claim 76 in which said dynamic dataset represents set-top box data.

35 78. The method of Claim 77, in which said static dataset represents demographic data.

79. The method of Claim 78 in which said unidentified correlation dataset values represent non-sampled demographic specifications.

80. A method of invalidating set-top box events, comprising the steps of:

monitoring set-top box events;

5 storing such events in an array;

calculating trends in such events;

invalidating set-top box events which deviate in a statistically significant manner from observed set-top box event trends, or which match previously defined invalid set-top box events;

10 placing such invalidated set-top box events in an array; and

calculating trends in such invalidated set-top box events such that some long-term trends may be revalidated, and to identify new set-top box event categories to be ignored.

add B3